Ultrasound Obstet Gynecol 2014; **43**: 437–443
Published online 12 March 2014 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.13198

Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology

M. G. TOLSGAARD*, C. RINGSTED†, E. DREISLER‡, A. KLEMMENSEN‡, A. LOFT§, J. L. SORENSEN*, B. OTTESEN* and A. TABOR*

**Department of Obstetrics, Juliane Marie Centre, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; †Department of Anesthesia and The Wilson Centre, University of Toronto and University Health Network, Toronto, ON, Canada; ‡Department of Gynaecology, Rigshospitalet, Copenhagen, Denmark; §The Fertility Clinic, Rigshospitalet, Copenhagen, Denmark*

KEYWORDS: learning curves; medical education; ultrasound assessment; ultrasound competence; ultrasound training

ABSTRACT

Objectives *To explore the reliability and validity of a recently developed instrument for assessment of ultrasound operator competence, the Objective Structured Assessment of Ultrasound Skills (OSAUS).*

Methods *Three groups of 10 doctors with different levels of ultrasound experience in obstetrics and gynecology were included. The novices had less than 1 month of experience, the intermediate group had 12–60 months of experience and the senior participants were all consultants. Fifteen participants performed transabdominal fetal biometry and the other 15 participants performed systematic transvaginal gynecological ultrasound scans. All scans were video-recorded and assessed by two blinded consultants using the OSAUS scale. The OSAUS scores were compared between the groups using the Kruskal–Wallis test, and pass/fail scores were determined using the contrasting-groups method of standard setting.*

Results *For the transabdominal fetal biometry examinations, the mean ± SD OSAUS scores of the novices, intermediates and senior participants were 1.5 ± 0.4, 3.3 ± 0.6 and 4.4 ± 0.4, respectively (P = 0.003). For the systematic transvaginal scans, the mean ± SD OSAUS scores of the novices, intermediates and senior participants were 1.8 ± 0.2, 3.1 ± 0.1 and 3.9 ± 0.5, respectively (P = 0.003). Post-hoc comparisons showed significant differences between each of the groups for both types of scans. The pass/fail score was 2.5 for the transvaginal scan and 3.0 for the transabdominal biometry examinations. The inter-rater reliability was 0.89.*

Conclusions *Ultrasound competence can be assessed in a reliable and valid way using the OSAUS scale. The pass/fail scores may be used to help determine when trainees are qualified for independent practice.*

INTRODUCTION

Ultrasonography is considered safe but is highly operator dependent. The importance of ensuring adequate training of healthcare practitioners in the use of ultrasound has recently been highlighted in the National Health Service report on maternity claims, which identified ultrasound as one of the major areas for improving patient safety[1]. The international ultrasound societies recommend that trainees complete between 200 and 300 examinations before independent practice[2–4]; although this may overestimate the requirements for some trainees, it may be insufficient for others[5]. Moreover, there are large differences in the learning curves for different types of examinations as some may be learned quickly whereas others require repeated practice over longer periods of time[6,7]. These large variations underline the need for competency-based education rather than a one-size-fits-all approach to ultrasound training[5]. A central concept of competency-based education is to determine when trainees are sufficiently qualified for independent practice[8], therefore valid and reliable assessment instruments that can be used for certification, recertification and evaluation of training programs are needed.

To meet these demands, our group recently performed a consensus study involving ultrasound experts from multiple different specialties and countries to develop a scale for assessment of ultrasound competence, the Objective Structured Assessment of Ultrasound Skills (OSAUS)[5]. Development of OSAUS included several rounds of rating and commenting on what should be

Correspondence to: Dr M. G. Tolsgaard, Department of Obstetrics, Juliane Marie Centre, Copenhagen University Hospital Rigshospitalet, Blegdamsvej 9, DK-2100 OE Copenhagen, Denmark (e-mail: martintolsgaard@gmail.com)

Accepted: 19 August 2013

included in the assessment of ultrasound competence of trainees. Although expert consensus supports evidence of content validity of the assessment instrument[9], it is not known if the OSAUS scale actually discriminates between differences in trainee competence. The latter is known as 'construct validity', which is traditionally supported by differences in scores obtained from healthcare practitioners with increasing levels of competence[10]. Finally, before using a new assessment method instrument, pass/fail criteria should be established to determine when different trainees are fit for independent practice of different types of scans[11].

The aims of this study were therefore: (1) to determine if the OSAUS scale can be used to discriminate between trainees with increasing levels of ultrasound competence in obstetrics and gynecology; (2) to establish credible pass/fail scores of basic transabdominal and transvaginal ultrasound; and (3) to assess the reliability of the OSAUS scale.

METHODS

This was an experimental, rater-blinded study to determine construct validity and reliability of the OSAUS scale when used for transabdominal and transvaginal scans. Ethical approval was granted in the form of an exemption letter from the Regional Ethical Committee for the Capital Region (protocol no. H-1-2012-043). Figure 1 shows a flow chart of the study design.

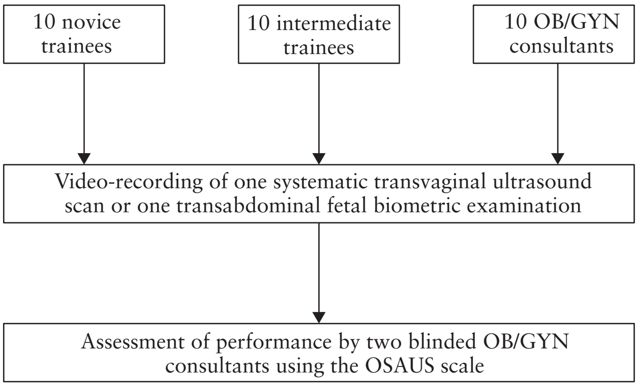
10 novice trainees
10 intermediate trainees
10 OB/GYN consultants
Video-recording of one systematic transvaginal ultrasound scan or one transabdominal fetal biometric examination
Assessment of performance by two blinded OB/GYN consultants using the OSAUS scale

Figure 1 Flow chart of study design. OB/GYN, obstetrician/gynecologist; OSAUS, Objective Structured Assessment of Ultrasound Skills.

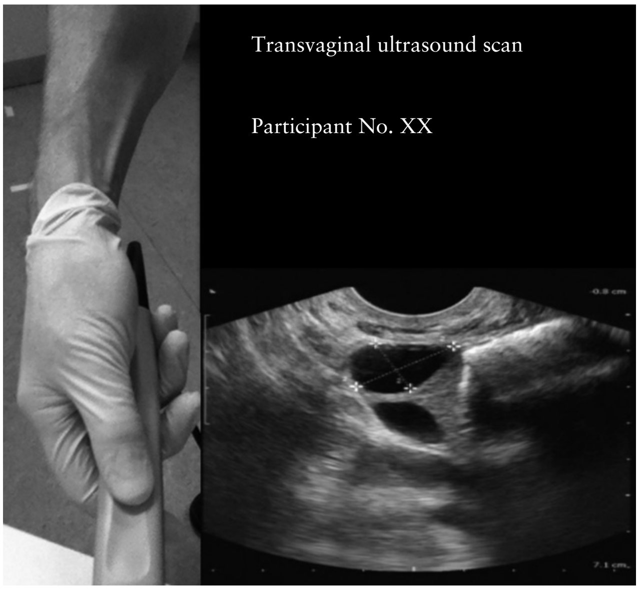

Figure 2 Still image from synchronized ultrasound output and video-recording. Audio distortion was added to the video clips to anonymize participants before two consultants assessed performance using the Objective Structured Assessment of Ultrasound Skills (OSAUS) scale.

The study was carried out at the Department of Obstetrics and The Fertility Clinic, Copenhagen University Hospital Rigshospitalet, Denmark. The transabdominal ultrasound examinations were performed during ambulatory consultations on pregnant women between 22 and 38 weeks of gestation with a preconception body mass index (BMI) of < 25 kg/m^2. Women undergoing transvaginal gynecological ultrasound examinations were fertility patients examined at The Fertility Clinic, Copenhagen University Hospital Rigshospitalet. None had received hormone stimulation at the time of the scan and all had a BMI of < 25 kg/m^2. All women provided informed consent before enrolling in the study. Hand movements were video-recorded and the outputs from the ultrasound devices were extracted using a hard-disk recorder. The ultrasound devices used in this study were GE Vivid E, GE Logic 700 MR (GE Healthcare Ultrasound, Milwaukee, WI, USA) and BK Medical Class I Type B (BK Medical, Peabody, MA, USA) with 3–5-MHz curved-array transabdominal probes and 5–9-MHz transvaginal probes. The ultrasound output and the video-recordings were synchronized and merged into a single video, and audio distortion was added to anonymize participants (Figure 2).

Participants included medical doctors with three different levels of experience of ultrasound in obstetrics and gynecology. The novices had less than 1 month of experience and no formalized ultrasound training, the intermediate group had 12–60 months of experience and the senior participants were all consultants in obstetrics and gynecology. The senior participants who performed transabdominal ultrasound were all fetal medicine consultants, and those who performed the transvaginal ultrasound scans were all fertility medicine consultants. All participants were recruited from the Department of Obstetrics or The Fertility Clinic, and all provided informed consent before participation.

The participants performing transabdominal ultrasound examinations were instructed to perform fetal biometry to obtain biparietal diameter (BPD), occipitofrontal diameter (OFD), anteroposterior abdominal diameter (APAD), transverse abdominal diameter (TAD) and femur length (FL). The participants performing transvaginal ultrasound examinations were instructed to perform a systematic pelvic ultrasound scan. A systematic scan included identifying and systematically scanning through the uterus in sagittal and transverse planes, measurement

Table 1 The Objective Structured Assessment of Ultrasound Skills (OSAUS), which included five items that were assessed on five-point Likert scales

	Likert scale		
Item	*1*	*3*	*5*
1. Applied knowledge of ultrasound equipment: familiarity with the equipment and its functions, e.g. selecting probe, using buttons and application of gel	Unable to operate equipment	Operates the equipment with some experience	Familiar with operating the equipment
2. Image optimization: consistently ensuring optimal image quality by adjusting gain, depth, focus, frequency, etc.	Fails to optimize images	Competent image optimization but not done consistently	Consistent optimization of images
3. Systematic examination: consistently displaying systematic approach to the examination and presentation of relevant structures according to guidelines	Unsystematic approach	Displays some systematic approach	Consistently displays a systematic approach
4. Interpretation of images: recognition of image pattern and interpretation of findings	Unable to interpret any findings	Does not consistently interpret findings correctly	Consistently interprets findings correctly
5. Documentation of examination: image recording and focused verbal/written documentation	Does not document any images	Documents most relevant images	Consistently documents relevant images

Likert is a five-point scale with 1 representing very poor and 5 representing excellent. In the OSAUS rating scale, only three points have descriptive anchors.

of the endometrial thickness, systematically examining the lateral pelvic wall, identifying and measuring both ovaries in three planes and examining the pouch of Douglas for free fluid. After completing the examinations, the participants were asked to document the examination in terms of a scan report. A clinical supervisor guided the novice group if they were unable to progress with the examination. This included technical assistance to operate the equipment and helping to identify key anatomical landmarks during the ultrasound examinations. Written information on the type of assistance provided was made available to the raters when assessing performance. There was an upper time limit of 20 min per examination.

The OSAUS scale consists of seven items: ‘indication for the examination’, ‘applied knowledge of ultrasound equipment’, ‘image optimization’, ‘systematic examination’, ‘interpretation of images’, ‘documentation of the examination’ and ‘medical decision-making’ (Table 1). In the original outline of the OSAUS scale, the first and final items were only to be included when appropriate. In this study, these two items were not included because the participants were specifically instructed to perform the examinations and because they were not necessarily the healthcare practitioners in charge of the medical management. Therefore, participants could not be assessed on their knowledge of the indications for the examination or on their decision-making skills and hence only the technical aspects of their performance were assessed. Each item was rated using the five-point Likert-scale provided, with descriptions of performance ranging from very poor (score = 1) to excellent (score = 5).

The raters were consultants who had a research background in fetal medicine and gynecological ultrasound. The raters completed comprehensive rater training before assessment of participant performance. The training consisted of rating videos of transabdominal and transvaginal ultrasound examinations, which were obtained using the same methods described above. During this rater training, four videos were first rated individually and then discussed until agreement between the raters was obtained. They were instructed to rate according to the standard expected from a trainee who had just completed specialty training in obstetrics/gynecology.

Statistical analysis

The mean ± SD were calculated for each group of participants for the transabdominal and transvaginal ultrasound examinations, respectively. The mean OSAUS scores were compared between the three groups using the Kruskal–Wallis test, and planned post-hoc comparisons were made between novices and intermediates and between intermediates and consultants using the Mann–Whitney *U*-test. Descriptive analysis was performed for the five OSAUS items across the three groups. Differences between the performances of fertility medicine consultants and fetal medicine consultants were then compared for each of the five OSAUS items.

Pass/fail scores were determined for each type of ultrasound examination using the contrasting-groups method[12]. Using this method, a group of non-competent performers (the novice trainees) was compared with a group of competent performers (the consultants) to establish the best discrimination between these two groups. The pass/fail scores were determined as the intersection between the distributions of OSAUS scores obtained from the novice trainees and the consultants. This was performed to give equal weight to non-competent performers who passed (false positives) and competent performers who failed (false negatives)[11].

Internal consistency was calculated using Cronbach's alpha, and inter-rater reliability was determined using intra-class correlation coefficients (ICCs) (absolute agreement single-measures). All analyses were performed using SPSS v. 20 (IBM Corporation, Arnock, NY, USA).

RESULTS

All 30 participants completed one ultrasound examination on one of 30 different women. The median (range) experience level for the novices, intermediates and fetal medicine consultants performing transabdominal ultrasound were 1 (0–3) weeks, 25 (13–60) months and 20 (10–23) years, respectively. The median (range) experience level for the novices, intermediates and fertility medicine consultants performing transvaginal ultrasound were 1 (0–4) weeks, 12 (12–15) months and 12 (10–30) years, respectively. For the fetal biometry examinations, the mean OSAUS scores of the novices, intermediates and senior participants were 1.5 ± 0.4, 3.3 ± 0.6 and 4.4 ± 0.4, respectively ($P = 0.003$). For the systematic transvaginal scans, the mean OSAUS scores of the novices, intermediates and senior participants were 1.8 ± 0.2, 3.1 ± 0.1 and 3.9 ± 0.5, respectively ($P = 0.003$). Post-hoc comparisons showed significant differences between novices and intermediates ($P = 0.009$) and between intermediates and seniors ($P = 0.028$) for the transabdominal scans. Similarly, post-hoc comparisons showed significant differences between transvaginal scans performed by novices and intermediates ($P = 0.009$) and between intermediates and seniors ($P = 0.035$). The distribution of mean scores on the five OSAUS items in the three groups is shown for each type of examination in Figure 3. The fetal medicine consultants scored significantly higher than the fertility medicine consultants on the item 'Image optimization': (mean ± SD) 4.1 ± 0.7 *vs* 2.5 ± 0.6, $P = 0.014$. There were no other significant differences between item scores of the two groups of consultants. The time used for the fetal biometry examinations was 9 min 55 s (SD = 2.6 min), 13 min 40 s (SD = 2.1 min) and 7 min 53 s (SD = 2.1 min) for the novices, intermediates and seniors, respectively ($P = 0.018$). Correspondingly, the use of time for the transvaginal scans was 9 min 42 s (SD = 4.5 min), 9 min 47 s (SD = 3.6 min) and 5 min 28 s (SD = 2.7 min) for the novices, intermediates and seniors, respectively ($P = 0.11$). The use of time correlated poorly to OSAUS scores (Pearson correlation coefficient = 0.38

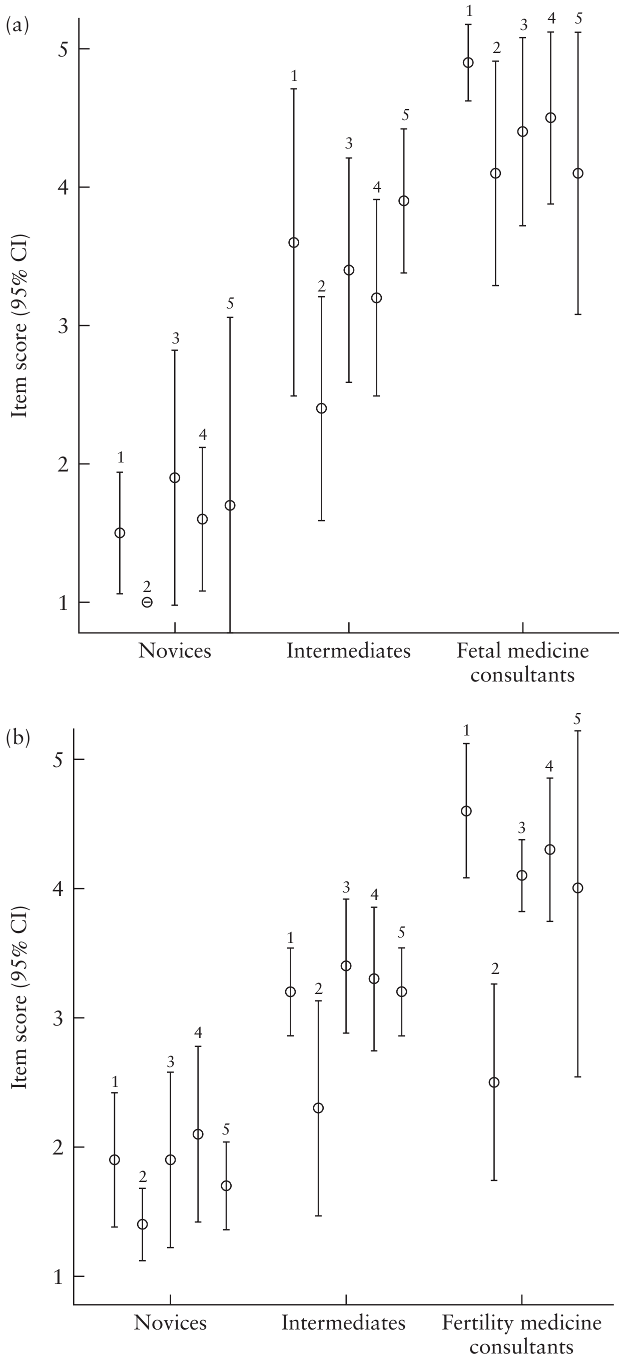

Figure 3 Distribution of scores on the five Objective Structured Assessment of Ultrasound Skills (OSAUS) elements for the transabdominal fetal biometry (a) and the systematic transvaginal ultrasound (b) examinations. Item scores: 1, applied knowledge; 2, image optimization; 3, systematic examination; 4, interpretation of images; 5, documentation of examination. The item 'image optimization' was scored significantly lower in the transvaginal scans performed by fertility medicine consultants than it was in the transabdominal scans performed by the fetal medicine consultants ($P = 0.014$).

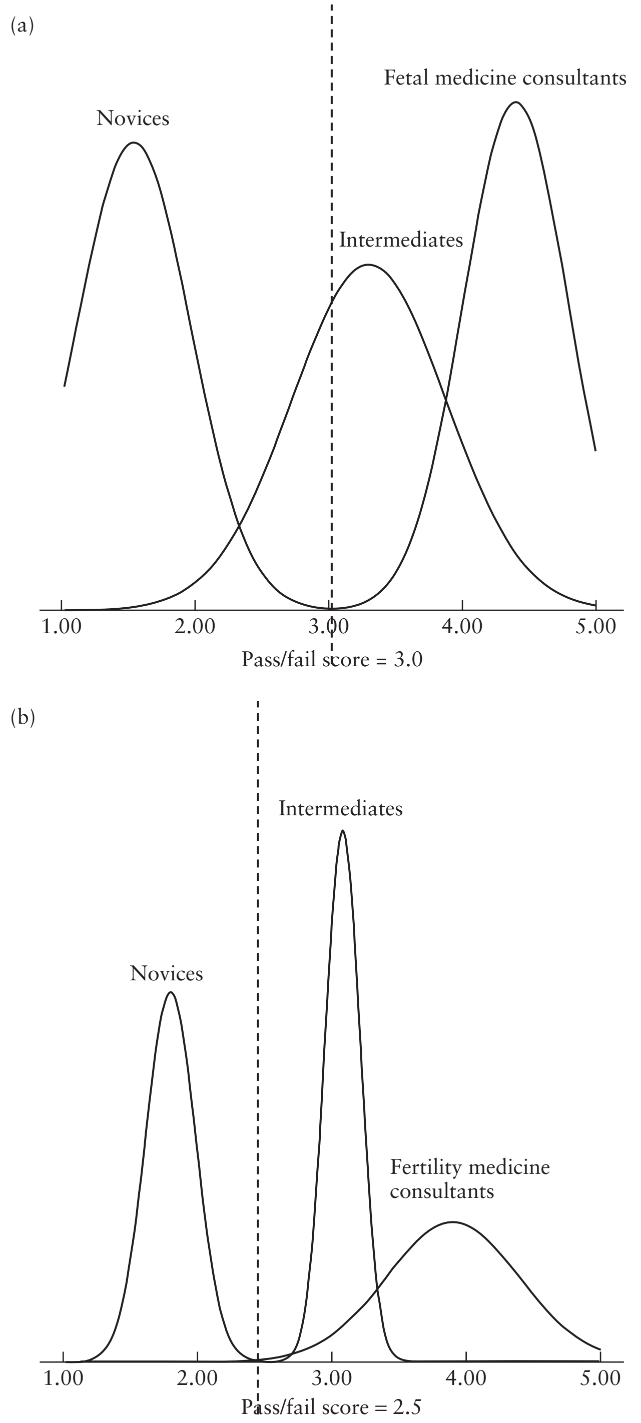

Figure 4 Pass/fail scores for the transabdominal fetal biometry examinations (a) and the transvaginal systematic ultrasound scans (b) using the contrasting-groups method. Using these criteria, no non-competent performers passed (i.e. there were no false positives) and no competent performers failed (i.e. there were no false negatives).

for the transvaginal scans and 0.20 for the fetal biometry examinations; $P = 0.16$ and $P = 0.48$, respectively). There were no significant differences in the gestational age of women being scanned in the three groups ($P = 0.53$) and the OSAUS scores were not significantly correlated to gestational age (Pearson correlation coefficient $= -0.32$; $P = 0.24$).

Pass/fail scores were determined at OSAUS mean 3.0 for transabdominal ultrasound and OSAUS mean 2.5 for transvaginal ultrasound (Figure 4). Using these criteria, all novices failed (there were no false positives) and all senior clinicians passed both types of examination (there were no false negatives). All intermediates passed the transvaginal examinations, but two out of five failed the transabdominal examinations. Both inter-rater reliability and internal consistency were high (ICC $= 0.89$; Cronbach's alpha $= 0.96$). There were no missing data.

DISCUSSION

This study demonstrates that the OSAUS scale discriminated between trainees with different levels of competence and established credible pass/fail scores for two types of ultrasound examinations. There was high internal consistency, as well as a very high inter-rater reliability, suggesting good agreement between raters. Hence, this study addresses recent calls for valid and reliable assessment instruments to ensure quality of the scans provided by trainees in obstetrics and gynecology[1,5,13].

Pass/fail scores for both types of examination were determined to allow future meaningful use of the assessment instrument. These standards did not produce any false negatives (consultants who failed) nor false positives (novices who passed) but 40% of the intermediate trainees failed to reach pass levels when performing the transabdominal scans. However, all intermediate trainees who performed transvaginal ultrasound reached the pass level, although their mean experience in obstetrics and gynecology was lower than that of the trainees who performed transabdominal ultrasound. These differences suggest that proficiency is attained more rapidly for the basic gynecological transvaginal scan than for transabdominal fetal biometry examination. This may be a consequence of what the trainees are taught and what is expected from them at different levels of training. In our setting, junior trainees are expected to manage early-pregnancy scans independently after just a few months of training, whereas proficiency in transabdominal fetal biometry examination is not expected during the first year of training. The differences in pass rates may, in part, also be explained by the lower pass/fail scores for transvaginal ultrasound compared with transabdominal ultrasound as a consequence of the slightly lower scores of the fertility medicine consultants (mean $= 3.9$) compared with the fetal medicine consultants (mean $= 4.4$). Furthermore, it is unlikely that the same pass/fail scores would be obtained in two different types of ultrasound scan with different focus of interest and type of equipment.

All items of the OSAUS scale were weighted equally in the tradition of existing assessment instruments used in other areas of medicine[11,14–17]. This approach was chosen because weighting of different items would increase computation and thereby limit the usability of the OSAUS scale in a clinical context[9]. Furthermore, when items of a scale are highly correlated, several studies have demonstrated that item weighting has little or no effect on a scale's correlation to other variables, and hence its validity[18–22]. In our study, the inter-item correlation was very high (Cronbach's alpha = 0.96) and we therefore believe that the use of unweighted items is justified. In other words, trainees who scored high on one item also scored high on all the other items, and similarly for those who scored low[9].

Previous studies on ultrasound learning have focused on the number of examinations needed for competence but this may be an unreliable measure of competence as a result of the large individual variations in learning curves[6,7,23]. A one-size-fits-all approach to ultrasound may result in insufficient skills of some trainees whilst overestimating the training requirement of others[5]. This is accounted for in competency-based education but, until recently, developing and validating methods for assessment of ultrasound competence had not received much attention. Nonetheless, competency-based ultrasound education is needed to avoid suboptimal quality of care by healthcare professionals who have not yet gained the knowledge and skills needed to practice independently[24] and therefore it should be considered an ethical imperative to ensure proficiency of trainees. Furthermore, the large financial costs associated with maternity claims[1] may, alone, justify allocation of resources to in-training assessment of trainees until they attain proficiency in performing independent ultrasound scans. Quality assurance has previously been addressed by introducing cumulative sum (CUSUM) scores in some centers to detect suboptimal performance by sonographers of routine scans[25,26]. However, this surveillance system can be labor intensive and relies on predefined metric or gold standards, making it difficult to use for assessment of trainee performance in large and dispersed units[27]. In this study, the use of time per scan was also recorded but it was not a valid measure of performance because it reflected neither different levels of experience nor competence. Other valid measures – such as OSAUS – are therefore needed to ensure quality of the scans performed outside specialized ultrasound units by trainees and consultants, who may never have completed structured training programs or undergone formal certification.

Although there is evidence of validity and high reliability of the OSAUS scale, there are also some limitations with regard to generalization. Validity is traditionally defined as the extent to which a test measures what it is purported to measure[10]. An assessment instrument can therefore never be valid *per se*, but only in relation to a specific purpose and only to a certain extent[28]. Consequently, the inferences made about the validity of the OSAUS scale are limited by the people, tasks and context of this study[9]. In terms of consequential validity, the pass/fail scores should therefore be re-evaluated before being used in other settings with different raters. Moreover, the pass/fail scores should not be interpreted as a single measure of when a trainee is fit for independent practice or as a measure of when training is complete. Rather, they should be considered as a benchmark of performance during training alongside other markers of competence, such as global ratings and diagnostic accuracy. Finally, this study is also limited by the format with which the assessments were performed. We used video- and ultrasound recordings to ensure rater blinding, but the intended use of the OSAUS scale is for in-training assessments of trainees during supervision of their performances, to enable educators to provide structured feedback to trainees based on their OSAUS ratings. The implications this has on learning, and thereby on the speed with which trainees attain proficiency in performing ultrasound scans independently, is a subject for future studies.

REFERENCES

1. NHS: Ten Years of Maternity Claims. http://www.nhsla.com/Safety/Documents/Ten%20Years%20of%20Maternity%20Claims%20-%20News%20Release%20-%20October%202012.pdf [Accessed 10 June 2013].
2. EFSUMB: European Federation of Societies for Ultrasound in Medicine and Biology http://www.efsumb.org/guidelines/guidelines01.asp [Accessed 10 June 2013].
3. AIUM: American Institute of Ultrasound in Medicine. http://www.aium.org/resources/statements.aspx. [Accessed 10 June 2013].
4. ISUOG: International Society of Ultrasound in Obstetrics and Gynecology training guidelines (http://www.isuog.org/StandardsAndGuidelines/Statements+and+Guidelines/Training+Guidelines/Default.htm) [Accessed 10 June 2013].
5. Tolsgaard MG, Todsen T, Sorensen JL, Ringsted C, Lorentzen T, Ottesen B, Tabor A. International multispecialty consensus on how to evaluate ultrasound competence: a delphi consensus survey. *PLoS One* 2013; **8**: e57687.
6. Jang TB, Ruggeri W, Dyne P, Kaji AH. Learning curve of emergency physicians using emergency bedside sonography for symptomatic first-trimester pregnancy. *J Ultrasound Med* 2010; **29**: 1423–1428.
7. Bazot M, Daraï E, Biau DJ, Ballester M, Dessolle L. Learning curve of transvaginal ultrasound for the diagnosis of endometriomas assessed by the cumulative summation test (LC-CUSUM). *Fertil Steril* 2011; **95**: 301–303.
8. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver I, Taber S, Talbot M, Harris KA. Competency-based medical education: theory to practice. *Med Teach* 2010; **32**: 638–645.
9. Streiner DL, Norman G. Validity. In *Health Measurement Scales: a Practical Guide to their Development and Use*. Oxford Medical Publications: Oxford, UK, 2008; 247–274.
10. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, 1999.
11. Konge L, Annema J, Clementsen P, Minddal V, Vilmann P, Ringsted C. Using virtual-reality simulation to assess performance in endobronchial ultrasound. *Respiration* 2013; **86**: 59–65.
12. Livingston S, Zieky M. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton: Educational Testing Service, 1982.
13. Salvesen KA, Lees C, Tutschek B. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? *Ultrasound Obstet Gynecol* 2010; **36**: 525–529.
14. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of

technical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84**: 273–278.

15. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003; **138**: 476–481.
16. Maagaard M, Oestergaard J, Johansen M, Andersen LL, Ringsted C, Ottesen B, Sørensen JL. Vacuum extraction: development and test of a procedure-specific rating scale. *Acta Obstet Gynecol Scand* 2012; **91**: 1453–1459.
17. Tolsgaard MG, Arendrup H, Lindhardt BO, Hillingsø JG, Stoltenberg M, Ringsted C. Construct validity of the reporter-interpreter-manager-educator structure for assessing students' patient encounter skills. *Acad Med* 2012; **87**: 799–806.
18. Lei H, Skinner HA. A psychometric study of life events and social readjustment. *J Psychosom Res* 1980; **24**: 57–65.
19. Streiner DL, Norman GR, McFarlane AH, Roy RG. Quality of life events and their relationship to strain. *Schizophr Bull* 1981; **7**: 34–42.
20. Gulliksen HO. *Theory of Mental Tests*. New York, NY: Wiley, 1950.
21. Streiner D, Miller HR. The MCMI-II: How much better than the MCMI? *J Pers Assess* 1989; **53**: 81–84.
22. Retzlaff PD, Sheehan EP, Lorr M. MCMI-II scoring: weighted and unweighted algorithms. *J Pers Assess* 1990; **55**: 219–223.
23. Vayssière C, Morinière C, Camus E, Le Strat Y, Poty L, Fermanian J, Ville Y. Measuring cervical length with ultrasound: evaluation of the procedures and duration of a learning method. *Ultrasound Obstet Gynecol* 2002; **20**: 575–579.
24. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 2010; **376**: 1923–1958.
25. Weerasinghe S, Mirghani H, Revel A, Abu-Zidan FM. Cumulative sum (CUSUM) analysis in the assessment of trainee competence in fetal biometry measurement. *Ultrasound Obstet Gynecol* 2006; **28**: 199–203.
26. Biau DJ, Porcher R, Salomon LJ. CUSUM: a tool for ongoing assessment of performance. *Ultrasound Obstet Gynecol* 2008; **31**: 252–255.
27. Nitsche JF, Brost BC. Obstetric ultrasound simulation. *Semin Perinatol* 2013; **37**: 199–204.
28. Shuwirth L, Van Der Vleuten C. Assessing competence: extending the approaches to reliability. In *The Question of Competence: Reconsidering Medical Education in the Twenty-First Century*. Hodges B, Lindgard L (eds). Cornell University Press: Ithaca, New York, 2012; 113–130.